



TITLE:

北大の実験と可能性：メタデータ交換  
プロトコルOAI-PMHに準拠した  
e-printサーバ構築 (電子情報交換に  
関する最近の話題)

AUTHOR(S):

行木, 孝夫; 畠山, 元彦

---

CITATION:

行木, 孝夫 ...[et al]. 北大の実験と可能性：メタデータ交換プロトコルOAI-PMHに準拠した  
e-printサーバ構築 (電子情報交換に関する最近の話題). 数理解析研究所講究録 2005,  
1446: 40-57

ISSUE DATE:

2005-08

URL:

<http://hdl.handle.net/2433/47648>

RIGHT:

# 北大の実験と可能性 — メタデータ交換プロトコル OAI-PMH に準拠した e-print サーバ構築\*

NAMIKI Takao, HATAKEYAMA Motohiko

Department of Mathematics,

Hokkaido University

行木孝夫<sup>†</sup>、畠山元彦

北海道大学大学院理学研究科数学専攻

## 概要

北海道大学数学教室における e-print サーバの構築について述べ、利用するソフトウェアとプロトコルを紹介する。メタデータを OAI-PMH プロトコルに載せて共有することで情報を有効に活用できることを示し、応用例として収集したメタデータの解析を示す。

## 1 はじめに

近年のネットワーク環境の変化は著しく、研究に関する主な情報交換はメーリングリストと web とへのシフトを完了しつつあると言っても過言ではない。しかしながら、利便性を向上させたかに見える ML と web とを中心とした情報交換には新たな問題点が指摘されている。あらかじめ情報の所在を知らなければアクセスできないという問題である。

以前は数学教室の掲示板に貼られていた情報が電子的に流通する結果、その情報を研究グループを核とするコミュニティの外から得るにはサーチエンジンの提供する雑多な情報を濾過しなければならない。

この問題をある程度まで解決するには、何らかの形で情報の所在を示すメタ情報を共有する機構が必要であり、REST アーキテクチャの一種である OAI-PMH によって実装することが可能になると考えられる。同様の発想は国立情報学研究所の JuNii [9] に見られる。

OAI-PMH とは The Open Archives Initiative Protocol for Metadata Harvesting の略である。対応したサーバはメタデータを全て (あるいはクライアントの指定した条件に応じて) 提供する。クライアントの指定しうる条件はサーバ側が特定のリクエストに応じて提供する。日付指定による差分の提供は独立したリクエストとして定義されている。

同時に、数学教室における e-print サーバの構築については多くの利点がある反面、いくつかの問題点が指摘されている。日本において必ずしも e-print サーバが普及していない理由の一つには、

\*本研究は 21 世紀 COE プログラム「非線形と特異性の数学」情報文献機能におけるプロジェクト「数学の海」の一環である。

<sup>†</sup>nami@math.sci.hokudai.ac.jp

管理上の問題とは別に、電子化への慎重な立場から指摘される「盗作」への対処が難しい点である。この完全な解決は困難だが、北大数学教室では e-print サーバ間の連携に OAI-PMH を採用するサーバプログラムを利用することで、問題の解決を図りながら e-print サーバの構築を行った。

本稿に述べた e-print サーバは 2003 年 11 月から試行を始め、2004 年 1 月から実運用に入っている。以下では OAI-PMH を実装するソフトウェアの選定から解説する。

## 2 概要

OAI-PMH に対応したメタデータサーバをリポジトリ、データプロバイダとよぶ。分散したデータプロバイダから OAI-PMH によってメタデータを収集し、何らかの機能を提供するサーバをサービスプロバイダとよぶ。ユーザはサービスプロバイダの提供する機能を利用する (図 1)。

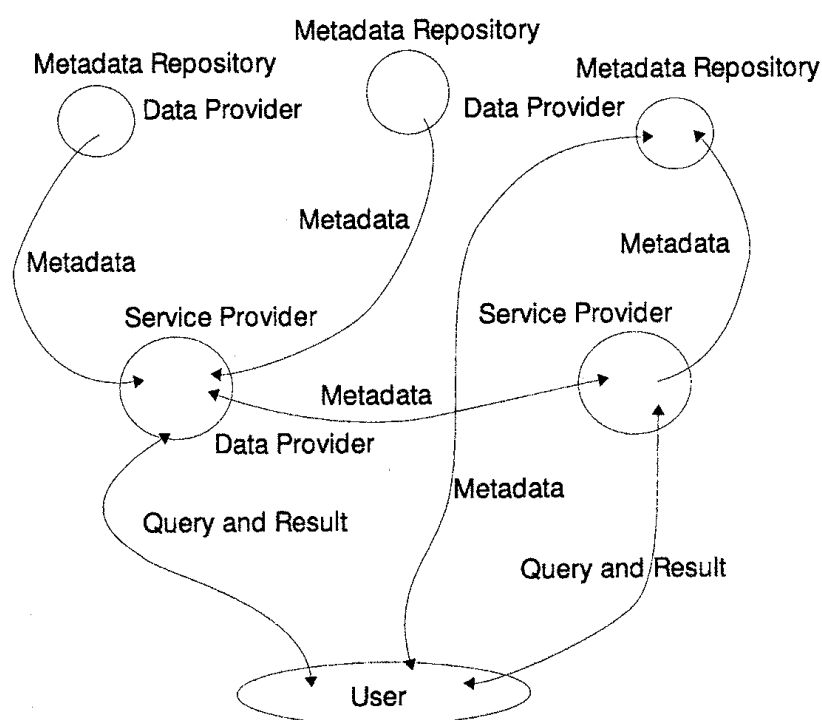


図 1: 概念図

データプロバイダの構築とサービスプロバイダの構築とは分けて考える。データプロバイダの構築自体は後述するソフトウェアを利用すればよい。以下ではデータプロバイダの構築とサービスプロバイダの構築とを紹介する。

### 3 データプロバイダの構築

#### 3.1 ソフトウェアの選定

E-print サーバの実現に際しては、適切な形式でファイルを置き、目次としての HTML を書くだけのサーバであれば簡単である。しかし、投稿者管理、版管理、OAI-PMH の実装などを考えると相当の規模になることを考慮し、既存のオープンソースソフトウェアを利用することにした。候補は表 1 の二件であり、それぞれに特徴がある。

ソフトウェア	所在	開発元	特徴
GNU EPrints	<a href="http://www.eprints.org">www.eprints.org</a>	Southampton University	Perl ベース。apache+mod_perl+MySQL
DSpace	<a href="http://www.dspace.org">www.dspace.org</a>	MIT と HP の共同開発	Java ベース。JSP+PostgreSQL

表 1: ソフトウェアの候補

どちらも必要な機能を満たしている。機能比較は [8] などを参照のこと。両者を比較すれば DSpace は大規模なサーバに向いている。「大規模」とは、複数の電子ジャーナルを同時に構築するという場合、あるいは複数の学部にもたがるリポジトリを構築する場合である。

数学教室の規模を勘案し、Perl ベースの GNU EPrints を採用した。2003 年秋の版は 2.2.1 であり、これを利用しているが、現在の最新版は 2.3.6 である。若干の機能追加がある。海外での GNU EPrints の事例については [7] などを参照。

EPrints のスタートページを図 2 に示す。

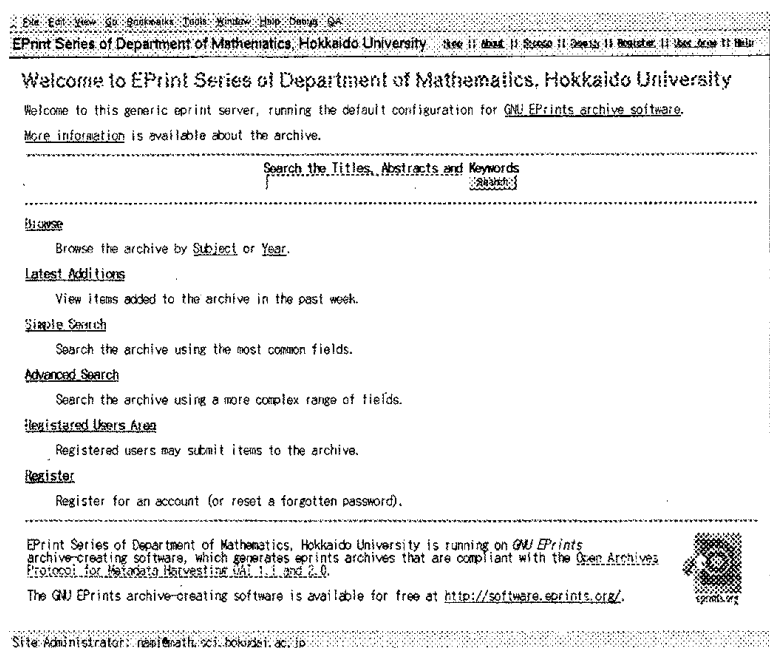


図 2: EPrints のスタートページ

### 3.2 コンテンツと分類

21 世紀 COE プログラム「非線型と特異性の数学」におけるサブプロジェクト「数学の海」のベースとして運用するため、収録するコンテンツはプレプリントに限定しない。

- 北大数学教室プレプリントシリーズ
- 北大数学教室講究録
- 研究集会の講演論文、予稿、論文集
- ソフトウェア
- ビデオアーカイブ (主に公開講座、一部の講義)
- Hokkaido Mathematical Journal(予定)

プレプリント、講究録は北大数学教室発行のものに限っている。教室メンバー、COE 協力研究者であれば任意に投稿できる。論文誌 Hokkaido Mathematical Journal は収録準備中であり、年度内には開始できると思われる。

研究集会の講演論文、予稿、論文集は主催者から申請があれば収録する。研究集会の担当者が講演者から集約し、まとめて登録する体制を整えた。本来は講演者に入力してもらうべきではあるが、一度きりの登録には入力が煩雑であろう。

ソフトウェア、ビデオアーカイブとは、数学関連ソフトウェア、講演記録を意味する。前者は無視できない存在であり、後者は近年の研究交流においては不可欠になるであろう。現在、ソフトウェアとしては KNOPPIX/Math を収録し、ビデオアーカイブとしては 2004 年の高校生向け公開講座を収録した。

コンテンツ指定部分を図 3 に示す。

### 3.3 主題リスト

主題リストには日本の数学者が最も慣れていると思われる American Mathematical Society の Mathematics Subject Classification 2000 [1] を採用した。一定水準の細かさがあれば、後に変換する必要が生じた場合にも柔軟に対応できると考えての選択である。現状では最も詳細な主題リストと考えられる。

MSC は数字 2 桁アルファベット数字 2 桁という形式であり、数学の各分野と周辺領域を網羅している。簡易化のために上位 2 桁を用いているが、この妥当性は今後の運用によって判明するはずのものである。セルフアーカイブによらない場合、事務担当者による投稿になるが、この場合は 00-xx とする。

Harvest の際に NDC へ対応するように変換表を作成した。原理的には LCSH などへの対応も可能である。

メタデータ登録の際に主題リストから選択する。その一部を図 4 に示す。

File Edit View Go Bookmarks Tools Window Help Debug QA

EPrint Series of Department of Mathematics, Hokkaido University    Home || About || Browse || Search || Register || User Area || Help

### Deposit Type

**EPrint Type \***  
Please select the most appropriate type for your deposit.

Book  
Book Chapter  
Conference Proceedings  
Conference Paper  
In Collection  
Conference Poster  
Technical Report  
Journal (On-line/Unpaginated)  
Journal (Paginated)  
Newspaper/Magazine Article  
Other  
Software  
Video  
Preprint  
Thesis

Cancel    Next >

---

Site Administrator: [nami@math.sci.hokudai.ac.jp](mailto:nami@math.sci.hokudai.ac.jp)

図 3: コンテンツ指定ページ

File Edit View Go Bookmarks Tools Window Help Debug QA

EPrint Series of Department of Mathematics, Hokkaido University    Home || About || Browse || Search || Register || User Area || Help

### Bibliographic Information

Please enter the bibliographic data about your deposit. Fields marked with a \* are fields that must be filled out before your deposit will be accepted.

**Authors \***  
Please enter the authors below. If there are more authors than available spaces, click on the "More Spaces" button. To remove an author, just remove their surname from the surname box.

1. Given Name(s)/Initials    Family Name(s)  
Person ID: .....

2. Given Name(s)/Initials    Family Name(s)  
Person ID: .....

3. Given Name(s)/Initials    Family Name(s)  
Person ID: .....

[More Spaces](#)

**Title \***  
Please enter the full title of the deposit.

.....

**Subjects v**  
Please select at least one main subject category, and optionally up to two other subject categories you think are appropriate for your deposit. In the list below, to select more than one subject, click on the "More" button, or SHIFT to select more than one subject.

00-xx GENERAL  
01-xx HISTORY AND BIOGRAPHY  
03-xx MATHEMATICAL LOGIC AND FOUNDATIONS  
05-xx COMBINATORICS  
06-xx ORDER, LATTICES, ORDERED ALGEBRAIC STRUCTURES  
08-xx GENERAL ALGEBRAIC SYSTEMS  
11-xx NUMBER THEORY  
12-xx FIELD THEORY AND POLYNOMIALS  
13-xx COMMUTATIVE RINGS AND ALGEBRAS  
14-xx ALGEBRAIC GEOMETRY

図 4: EPrints のメタデータ登録ページの一部。下方に主題リストの選択ボックス。

### 3.4 投稿者

セルフアーカイブを基本方針とした。投稿者は教室構成員とし、学外の COE 協力研究員を含む。問題が生じた時点で対応することになっている。現在の登録者は 17 名である。

ユーザ登録は学外から誰でも可能であり、投稿後に資格審査を行っている。2005 年 1 月には無資格で似非科学系のプレプリントを大量に投稿されたため、上記の審査を明文化した。これは、ある程度セルフアーカイブが浸透していることの証明とも受け取れる。

### 3.5 資料収集

#### 3.5.1 プレプリント、講究録

プレプリントシリーズは 2003 年秋発行の 600 号から収録し、年間 100 号のペースである。これは海外の数学教室発行のプレプリントシリーズとの交換という形で公式に配布している。講究録は研究集会の報告集が多く、著者の了解を取りきれない場合が多いため一部の収録にとどまっている。これも国内外の数学教室に配布している。

プレプリントの 599 号以前はスキャン済みであり、順次入力を進めている。

#### 3.5.2 Hokkaido Mathematical Journal

HMJ は有力なコンテンツであり、2004 年までに発行の第 33 巻までは既にスキャンを終え、公開準備にかかっている。2005 年刊行の第 34 巻より本文を含めた電子化が決定し、少なくともアブストラクトについては遡及して公開することになる。現在、第 34 巻 1 号と 2 号が公開されている。

EPrints には本文の公開を選択するオプションを付加でき、Apache のアクセス制御機構を利用して IP アドレスによるアクセス制限を付した。交換先へは IP アドレスによる公開である。このため、サーバを別に構築した。

#### 3.5.3 論文集他

一人のスタッフからは定期的に開催されている研究集会の報告集を収録したいという要請があった。代数幾何学城崎シンポジウムとして毎年開催されているもので、現在 85 年と 91 年をスキャンして収録した。院生の協力のもとに進行中である。主要メンバには了解をとったとのことである。

ファイルのアップロードには URI を指定できるので、既に web ページに載っている予稿集などの再収録には効率がよい。

#### 3.5.4 予稿収集の問題点

研究集会の予稿などは、参加登録と同時に収集すると効率が良い。しかし、高機能ゆえにやや煩雑な EPrints のインタフェースをそのまま使わせるのは無謀である。数ページの予稿を登録するためにユーザ登録と煩雑なアップロード作業はバランスが悪い。

一画面ですむ簡易登録フォームを用意した。後から一括変換を行なうことにする。これは EPrints の使う MySQL を直接操作することで可能になるが、まだ試みてはいない。

今後の課題として、認証を省き予稿登録に特化した EPrints を作りたい。他のツールで可能ならばそちらを使う。講究録は講義録や論文集が混じるため、EPrints とはなじまないかもしれない。講究録のみ DSpace にする可能性もある。

EPrints では、一件の登録につき複数のファイルを登録できる。そのため、論文集などの登録では一本ごとの登録にするか、まとめた登録にするか迷うことがある。特に講究録として刊行されるものは悩ましい。現在は余力があれば双方に登録することになっている。根本的な解決には付録 A に延べるように DSpace を使うことになろう。

### 3.5.5 セミナー、研究集会情報

プレプリント等の扱いはほぼ確立している。問題はセミナー、研究集会等に関するメタデータの持ち方である。北大では独自の CGI スクリプトによって入力を支援しているが、これを EPrints に移行して OAI-PMH に対応させる。このとき、特に研究集会については講演ごとにメタデータを作るかどうかなどの粒度が問題になる。

また、Dublin Core 以外に iCalendar 形式への対応も考えられる。これは既に幾何学分科会で実装している。

## 3.6 アーカイブの一覧表示

一般に、デジタル的なコンテンツを収集したサーバは検索機能の如何に関らず書庫に相当するものと考えられる。一方、目的のコンテンツを確定している利用者は稀であり、多くの利用者は一定の方針で整理された一覧表示を望むと思われる。これは開架に相当する。博物館であれば常設展示のようなものであろう。

適切に組織された一覧は、コンテンツが代表する分野の一般公開とも関連し重要である。

あるいは、書庫から一定の基準のもとに資料を選択して一時的に一覧を作成することも可能である。企画展に対応するものである。以下ではいくつかの研究集会の予稿集を教室の web ページに従って組版した例を示す。

EPrints 標準でもフィールドを指定して整列した一覧表示機能はある。本質的には資料タイプ、主題リスト、日付などによる分類による一覧である。

これによる機械的な一覧表示とは別に、教室の web ページとの統一的な外観を維持するため、OAI-PMH プロトコルの ListRecords リクエスト出力結果を加工し、HTML へ変換するツールを作成した。

資料タイプでの分類に留まらない表示、研究集会等での一覧が可能になった。

- 北大数学教室プレプリントシリーズ:

<http://coe.math.sci.hokudai.ac.jp/literature/preprint/back/preprint.aa.html>



- 代数幾何学城崎シンポジウム:  
<http://coe.math.sci.hokudai.ac.jp/literature/db/kinosaki.aa.html>
- 特異点理論とその応用:  
<http://coe.math.sci.hokudai.ac.jp/literature/db/sing.aa.html>

### 3.7 盗作、改変問題

#### 3.7.1 問題の所在

「盗作問題」は数学周辺において e-print を立ち上げる際、常に問題となる。プレプリントの著者欄だけを書き換えてそのまま投稿され、極めて稀ではあるが、査読の目をすり抜けることがある。場合によっては本物よりも早く出版されてしまう。電子化された場合には、それがさらに容易になるというのである。もちろん、そうそう生じることではないが、一度起きたことは強く印象づけられるという例であろう。プレプリントに記述したアイデアをあまり早期に広めたくないなどの理由から、ごく一部にだけ流通させたいという希望を持つ著者もいる。

専門誌への掲載以前には配布範囲を自分の眼の届く範囲に抑えたいということになる。決して少数派ではなく、正当性な主張でもあり、無視するわけにはいかない。解決のため、冊子体は従来通り作成し、電子版を希望しない場合には冊子体のみを作成することにした。

もう一つ、改変問題がある。プレプリントの役割は成果の先行を主張するものであるから、発行後に改変があっては信頼性に問題が出る。ここに神経を使う著者もいるが、それはシステムの移動等があっても本文になるべく触らないという手段を取らざるをえない。

#### 3.7.2 OAI-PMH による盗作、改変問題への対応

前述の通り、冊子体を残すことで問題はおおむね解決することになったが、盗作問題のみ残っている。電子化によって盗作が容易になるであろうことは明らかであり、これに解決の糸口を見つけねばならない。

問題は、電子化され、タイトルと著者を改変された情報から正しいプレプリントをレフェリーが見つけ出すのは困難であるという点に尽きる。OAI-PMH によって教室間のプレプリント情報を共有し、少なくとも abstract までを利用できることになれば、状況は大きく前進する。なぜなら、盗作者が abstract を正しく変更するという自体は考えにくいからである。

## 4 サービスプロバイダの構築

メタデータを取得するデータプロバイダとして、arXiv, EUCLID, NUMDAM, MathPreprints, JuNii を選択した。再利用を考慮し、ミラーサーバを設置する。

## 4.1 ミラーサーバの設置

ミラーサーバには [3] の関連ツールリストから oaii を選択した。これは ListRecords の出力を record ごとに 1 レコードとして RDB へ格納し、インターフェースとして CGI を用意している。以下にミラー状況を示す。

Identifier	URI	Harvest Method	Records
arXiv	www.arxiv.org	ListRecords	323549
euclid	projecteuclid.org	ListRecords	29541
MathPreprints	www.mathpreprints.com	ListRecords	1223
JUNII	ju.nii.ac.jp	ListRecords	67786
HUMATH	eprints.math.sci.hokudai.ac.jp	ListRecords	890
HMJ	hmj.math.sci.hokudai.ac.jp	ListRecords	1100
NUMDAM	www.numdam.org	ListRecords	8931

表 2: ミラーリング中のデータプロバイダとメタデータ件数

明らかに arXiv の量は頭抜けており、EUCLID がこれに続く。両者が数学関連の最大のデータプロバイダであることは間違いない。EUCLID は電子ジャーナルを供給し、arXiv はプレプリントであるから、重複も比較的少ないと考えられる。

ハーベストしたメタデータ数から見た arXiv の分野別投稿数を図 5 に示す。

他にも CogPrints (cogprints.ecs.soton.ac.uk) という 2 千件程度のプレプリントや論文を集めたデータプロバイダがある。認知科学関連の文献が主であるため、扱わなかった。CogPrints と arXiv からサービスプロバイダを構築した例が [11] に見られる。

## 4.2 サービスの提供

ミラーサーバへ定期的に ListRecords を発行し、検索可能な設定を行った [12]。ListRecords の出力を record ごとに分割し namazu にかけることで日本語へ対応している。また、補助的にミラーサーバの RDB へ直接 query をかけるスクリプトも用意した。

検索以外のサービスについては次の節を参照してほしい。

## 5 Mathematics Subject Classification の共相関解析

数学の研究において活発な分野を主観によらず取り出す手段はないだろうか。

計量書誌学において重要な要素として共引用、共著関係などがあり、それをもとに重要文献を抜き出すという手法が確立している。単純にインパクトファクターや被引用数を使うよりももっともらしい結果を得るが、その理由を示した文献は (少なくとも日本には) 存在しない。

実際は共相関行列 (あるいは相互情報量行列) を導出し、それによる解析を行っていることに帰着されるため、統計的に有意な手法となっていることが観察される。



MathSci CD-ROM 等を活用すれば同様の解析も可能であろうが、ここではメタデータハーベスティングの応用例として、収集したメタデータから各分野の論文一覧を作成し、相関関係を図示することにした。使用したメタデータは Project EUCLID の提供する oai\_dc であり、そこから全ての論文と、日本の数学教室発行のジャーナルとの二通りで行った。なお、後者には独自公開の Hokkaido Mathematical Journal も含めている。

解析は以下のように行った。

1. 各論文から指定されている AMS Mathematics Subject Classification(MSC) を抜き出し、各 MSC をノードとする有限グラフを作成する。エッジは一論文から同時に指定されている MSC について、部分完全グラフを生成することによって張っていく。
2. 上位二桁について粗視化する
3. 相関係数を  $N(X,Y)/\sqrt{N(X)N(Y)}$  によって決める。 $N(X,Y)$  はノード  $X,Y$  を結ぶエッジの数、 $N(X), N(Y)$  はそれぞれノード  $X,Y$  を指定した論文数である。

全 EUCLID に関しては簡明化のために相関係数の下位 1% を切り捨てている。1975 年からの可視化結果を [15] に示す。

このような手法は科学技術政策研究所発行の [13, 14] において Thomson ISI 社のデータベース ESI を用いた共引用解析に触発され、Thomson ISI では収録数が少ないと考えられる数学関連分野の再評価のために行うものでもある。

相関行列を可視化し図 6,7 に示す。詳細な解析は後に譲るとして、日本のジャーナルにおいては解析系と代数幾何系とが二つのクラスターを成しているという解釈が可能と思われる。EUCLID 全体との比較では、OR や Logic を含まないという特徴がある。EUCLID には応用系の雑誌が多いという特徴を忘れてはいけない。特に確率論の雑誌が多いため、60 に集中する傾向にある。

これらから、種々の数学的な解析が考えられる。図に示したものは別に、数理物理の専門誌である CMP と一般誌の Duke. Math. J. との比較も行った。

## 6 今後の展開

EPrints は e-print サーバとして開発が続いているソフトウェアであり、それに従う形での運用は既に述べたように比較的容易である。セミナー情報サーバ、メタデータサーバとしての運用は予定しているだけであるが、設定次第で可能であり、実現すれば総合的な研究情報サーバとしての見通しをつけられる。

OAI-PMH で提供するメタデータ形式を複数種類に対応させることで、MathNet [10] や JuNii [9] といったメタデータ関連プロジェクトとの相互運用を可能にできる。

最近のいわゆる blog の普及に伴い、メタデータは確実に普及し知名度を増している。その中で主役に近い役割を果たしているものは更新通知に用いられる RSS というメタデータの形式である。これには数種類の版があり議論の多い部分でもあるが、メタデータの主要部にはダブリンコアを採用している。

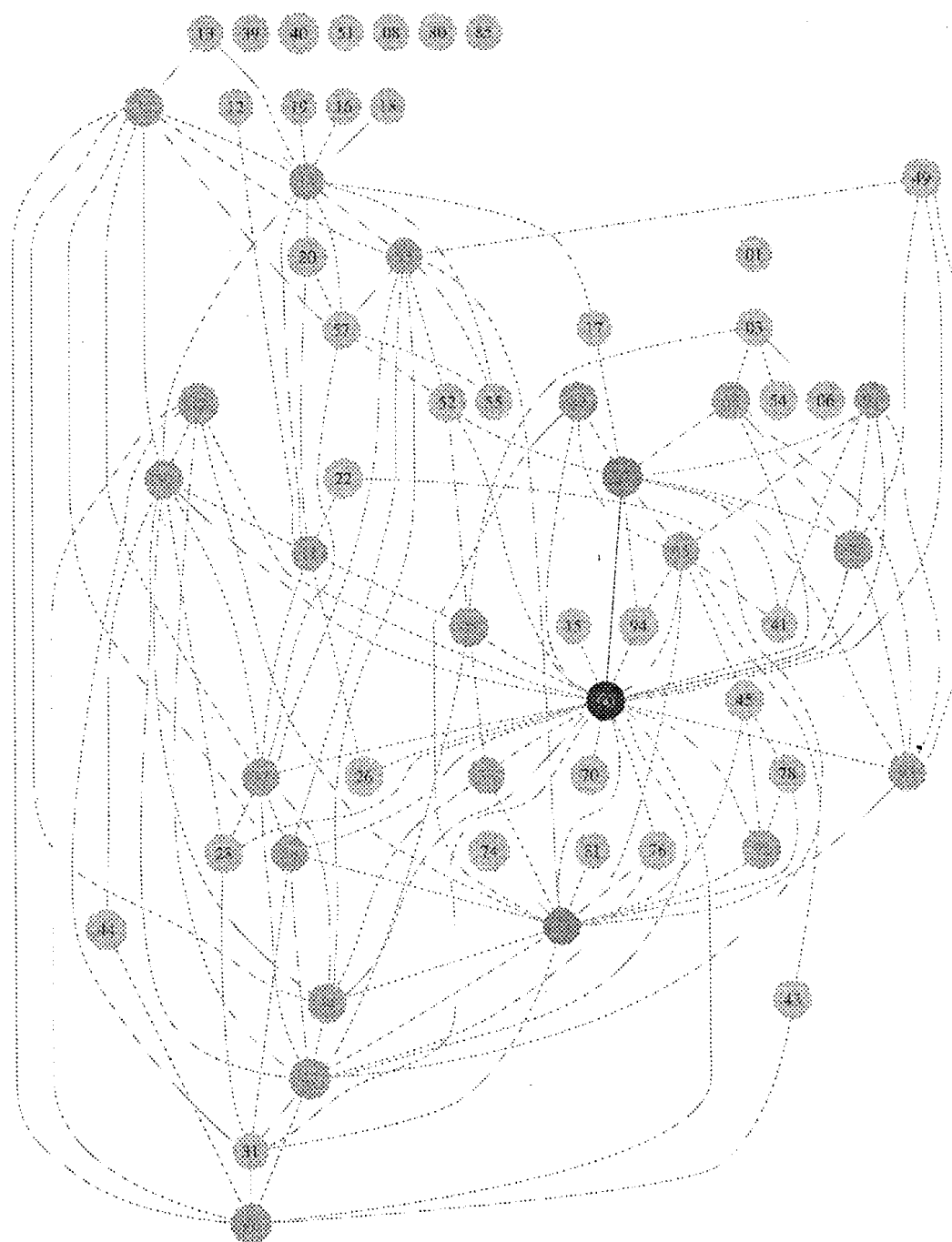


图 6: 2004 EUCLID

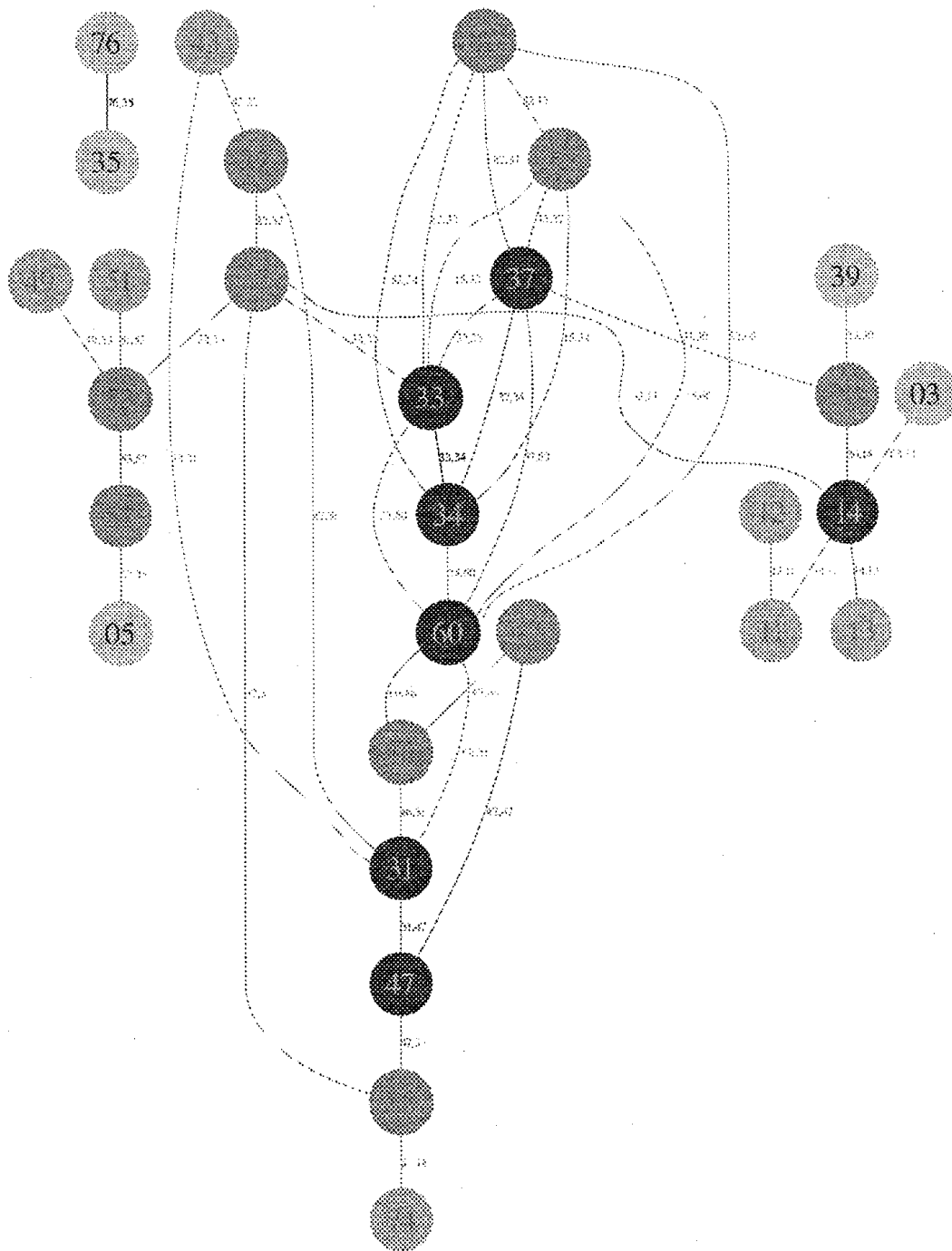


図 7: 2004 EUCLID から東北大、東工大、名古屋大

更なる展望として、更新通知を RSS で行いメタデータの完全な交換には OAI-PMH という方針が考えられる。

サービスプロバイダに関しては各所で検索以外のサービスを模索しているが、キーワードに関する自動更新などを考えたい。現在、ハーベストしたメタデータに関する検索と、メタデータと Google との同時検索を用意している。

2004 年 12 月現在、Google Schooler ベータ版は比較的良好な検索結果を返してくる。リポジトリとサーチエンジンとは相互に補完するものであると考えたい。

## A 機能面から見た DSpace と EPrints との比較

通常の web サーバを運営している環境であれば、mod\_perl ベースの EPrints は構造もわかりやすく、インストールから維持管理も web サーバ管理の延長として行える可能性がある。しかしながら、EPrints は資料タイプ別の管理を目的として構成されており、複数の資料タイプを混在させるコレクションを構築する場合には使い難い。従って、DSpace と EPrints の選択に当たっては規模以外にもコレクションの目的を考慮すべきである。

例を挙げる。北大数学教室の講義録は主に講義録と論文集から成り、後者には研究集会の予稿集が多い。従って、コレクションとしてはまず予稿集として一つのコレクションを形成し、それを講義録に収録するという形式が望ましい。これを EPrints で実現するのは難しい。

メタデータとしても、論文集としてのメタデータと収録各論文のメタデータとは区別し、双方を提供するべきであろう。

これは必ずしも DSpace を使うことを推奨するわけではない。資料タイプすなわちトップカテゴリとなるコレクションであれば簡素な構造の EPrints を採用することで管理コストを下げられる。要はコレクションの性質を見極めることである。

## B メタデータ交換プロトコル OAI-PMH の概観

### B.1 リクエストの概要

OAI-PMH は HTTP 上の GET メソッドへエンコードしたリクエストによって系統的にメタデータを取得するプロトコルである。CGI スクリプトによって実装されることが多い。最新版は 2.0 である。

リクエストは次の形式をとる。

```
http://server.domain/cgi-script?verb=request&request-options
```

主なリクエストを次に挙げる。

1. Identify: リポジトリの情報を取得。
2. ListSets: 選択しうる主題を取得。

3. ListMetadataFormats: メタデータ形式を取得。
4. ListRecords: 実際にメタデータを取得。オプション metadataPrefix は必須。

ListMetadataFormats によってメタデータ形式を取得した後、次のように ListRecords リクエストを発行すれば系統的にメタデータを取得できる。メタデータ形式 oai\_dc は必須であり、これを使うならば ListMetadataFormats の発行は必ずしも必要ではない。

```
http://server.domain/cgi-script?verb=ListRecords&metadataPrefix=oai_dc
```

## B.2 ListRecords によって取得されるメタデータの例

ListRecords を発行した後に返される XML を図 8 に示す。ListRecords タグ中に現れる record タグがメタデータ一件を示す。

## C メタデータの標準仕様 Dublin Core

### C.1 メタデータとは何か

メタデータとは、データに対する付加情報の総称を示す抽象的な概念である。今、一冊の書籍をデータとすると、次のようなメタデータが考えられる。

- 所在や目録情報 (URI, ISBN, 分類番号, 書店や図書館の書棚における位置など)
- 解説や注釈, 書評, 背景情報など
- 関連情報、参考文献、発展的な文献など

メタデータは多様な情報を扱うことになる。仕様が乱立すると相互利用に支障があるため、広く一般的に利用されられる情報に関しては共通の規格を用いることが多い。Dublin Core[2] は、そのような場合に用いられる仕様の一つである。RFC2413 に解説があり、RFC2731 には HTML の meta タグへのエンコード例が記載されている。特定分野においても MathNet[10] や OAI-PMH では標準として採用されている。

Web サイトの更新情報などを記述する RSS1.0(これもメタデータ的一种といえる)においてもサポートされており、最も普及した Dublin Core のアプリケーションの例であると思われる。

### C.2 Dublin Core の例

例として北大数学教室の web ページを示す Dublin Core メタデータの例を示す。

dc で Dublin Core を示し、コロンに続くキーワードは Dublin Core で機能が定められた token となる。identifier はメタデータの指示するものの所在 (図書館の蔵書であれば請



```

<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <responseDate >2004-08-13T23:26:19Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc" resumptionToken="">
    http://eprints.math.sci.hokudai.ac.jp/perl/oai2</request>
  <ListRecords>
  <record>
    <header>
      <identifier>oai:EPrintsHUMATH:0000009</identifier>
      <timestamp>2003-11-25</timestamp>
      <setSpec>7375626A656374733D33352D7878</setSpec></header>
    <metadata>
      <oai_dc:dc
        xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
          http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
        xmlns:dc="http://purl.org/dc/elements/1.1/">
        <dc:title>
          On viscous conservation laws with growing initial data
        </dc:title>
        <dc:creator>Yamada, Kazuyuki</dc:creator>
        <dc:subject>35-xx PARTIAL DIFFERENTIAL EQUATIONS</dc:subject>
        <dc:description>A unique local solvability is established for
          viscous conservation laws when the initial data may grow at
          the space infinity with a natural order. It is also show
          that such a classical solution can be extended to
          a global-in-time solution proved that the growth order of
          the initial data is less than critical order.
        </dc:description>
        <dc:date>2003-11-25</dc:date><dc:type>Preprint</dc:type>
        <dc:source>
          Preprint Series of Department of Mathematics,Hokkaido University
        </dc:source>
        <dc:identifier>
          http://eprints.math.sci.hokudai.ac.jp/archive/00000009/
        </dc:identifier>
        <dc:format>pdf</dc:format></oai_dc:dc></metadata></record>

```

図 8: ListRecords の出力例

dc:title	北海道大学数学教室
dc:identifier	http://www.math.sci.hokudai.ac.jp/
dc:type	web page
dc:format	text/html
dc:description	北大数学教室の top page

求番号など、あるいはISBNなど。e-print ならその URI など) を示し、type は data の分類を示す。format はデータ形式を表す。

HTML の meta タグを利用し、HTML ファイルとして実現すると図 9 のようになる。二重引用符の扱いには注意が必要である。

```
<html>
<head>
<meta name="dc.identifier" content="http://www.math.sci.hokudai.ac.jp/">
<meta name="dc.title" content="北海道大学数学教室">
<meta name="dc.type" content="web page">
<meta name="dc.format" content="text/html">
<meta name="dc.description" content="北大数学教室の top page">
</head>
</html>
```

図 9: HTML への meta タグによる実現

XML では図 10 のようになる。XML はタグの属する名前空間を指定することでタグの意味を規定するため、名前空間の指定には注意が必要である。

```
<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<dc:title>
  北海道大学数学教室
</dc:title>
<dc:type>Web page</dc:type>
<dc:format>text/html</dc:format>
<dc:identifier>http://www.math.sci.hokudai.ac.jp/</dc:identifier>
</oai_dc:dc>
```

図 10: XML での実現

## 参考文献

- [1] <http://www.ams.org/msc/>
- [2] <http://dublincore.org/>
- [3] Open Archives Initiative, <http://www.openarchives.org/>
- [4] <http://www.openarchives.org/OAI/openarchivesprotocol.htm>

- [5] <http://www.eprints.org/>
- [6] <http://www.dspace.org/>
- [7] Stephen Pinfield, Mike Gardner and John MacColl, *Setting up an institutional e-print archive*, Ariadne. Issue 31, March-April, 2002,  
(邦訳: <http://www.nii.ac.jp/metadata/oai-pmh/eprints/>)
- [8] Budapest Open Access Initiative, *A Guide to Institutional Repository Software*,  
<http://www.soros.org/openaccess/software/>  
(邦訳: [http://www.nii.ac.jp/metadata/oai-pmh/osi\\_ir\\_software\\_guide.html](http://www.nii.ac.jp/metadata/oai-pmh/osi_ir_software_guide.html))
- [9] <http://ju.nii.ac.jp>
- [10] <http://www.math-net.org>
- [11] <http://citebase.eprints.org>
- [12] <http://coe.math.sci.hokudai.ac.jp/literature>
- [13] NISTEP REPORT No.82 科学技術の中長期発展に係る俯瞰的予測調査 急速に発展しつつある研究領域調査 平成15年度調査報告書
- [14] NISTEP REPORT No.95 我が国における科学技術の状況と今後の発展の方向性
- [15] EUCLID 全体 <http://metron.math.sci.hokudai.ac.jp/search-new/navi>  
日本の数学教室発行雑誌 <http://metron.math.sci.hokudai.ac.jp/search-new/navi/jp>